AVIPE
Audio Visual IP Evaluator
And The Neuroscience of Network Traffic

This paper will discuss a new approach to scanning through large amounts of data. AVIPE, as will be defined below promises to be a paradigm shift in the way we compress big data and yet maintain intelligibility. This paper describes an innovative data processing solution that will improve current intelligence data analysis techniques, increase efficiency and extract new types of information from operationally relevant data. Our approach is to treat the overload of any source of data that confronts analysts from a speech processing point of view.

Historically, speech compression has been used to increase the transfer of information in the context of voice. This paper will set into place the novel approach of taking the point of view that most information operations (IO) can also be looked at in the same context as voice and be equated to a distinct set of dialect and idiolect components. ***Dialect refers to a language that is spoken within a group setting while idiolect refers to the features in the language of an individual.*** It will be shown that not only can the technique of human language analysis be applied to structured machine language but that more importantly this type of signal processing can be done in virtual real time. This tool will be easily integrated into existing platforms and can serve the user/analyst as both audible and visual aids. Our proposed Audio Visual Intelligent Protocol Evaluator (AVIPE) capability will be of great benefit to big data analysts in general by providing a method for screening real time intelligence network traffic and data flows for key elements of information that can then be backed out and analyzed for intelligence content. AVIPE will result in a new data processing paradigm that will increase efficiency by indicating areas that don't warrant further analysis while directing attention to those areas of information that do.

1.1 Overall Description.

The detection of signals, events, or patterns of interest is a continuing problem for the commercial and non-commercial communities which, has been made more difficult by the requirement that the detection must be done in real time. Word (or protocol) matching techniques while highly successful in off-line analysis are inadequate for on-line analysis. Efforts towards parallel processing and increasing processor speed to keep up with current demands are being outpaced by the growing amounts of information. It is generally recognized that matching techniques at the word level are lacking the requisite qualities needed to make a better informed and quicker decision.

The objective here is not to promote the development of a faster matching technique, but rather to offer a fundamental change and high level approach to pre-filter data in order to alert the operator or automated system analysis tool to apply a lower level matching tool. This fundamental change which involves the treatment of data passing through computer network

systems as a series of building blocks akin to the attributes associated with dialect and idiolect will bring about a dramatic increase in system analysis through-put, which of course, will come at a cost of resolution. However, it is projected that the 'higher altitude view', will yield patterns of interest, to which one may then intelligently pick which areas to drill down into, akin to a Hawk in the sky.

Cepstral analysis, a well understood tool in the context of voice/speech, is the technology vehicle of choice for AVIPE. Just like in the speech processing case where the cepstral components come about by grouping the underlying harmonics related to the glottal response, the components generated by AVIPE are derived from the grouping of underlying structure of subframes. Thus AVIPE can be thought of as a speaker ID-like coarse filter for data analysis across various intelligent protocol domains.

The advantages of incorporating this device, besides its ability to keep up with the data flow in virtual real time, includes the ability to partition the signal into homogeneous regions based on the cepstral coefficients that lay the foundation for the ensuing time domain and image processing analysis. The project is sensitive to operator control of the settings depending on the exploitation tasking at hand. The execution of the proposed algorithm will be carried out in this manner:

a) Store short term raw data.
b) Apply AVIPE through an optional GV^{TM}3.0 GUI to partition events/shapes of interest.
c) If applicable, make inquiry to other platforms.
d) Use raw data for final analysis.

The AVIPE Graphical User Interface (GUI) is shown in Figure 1. It will be built into PAR's viewer with the operator in mind. The data can be transformed several ways visually for independent views of the network traffic. The bottom right plot is of a framed session of network traffic in bit form. The remaining three plots are speech processing components in multiple forms. The components are the result of a 1000:1 reduction in streaming traffic to yield a (coarse filter) overview of changes that will alleviate the overload of data the operator has to contend with yet retain identifiable characteristics of the traffic flow. The operator will have the choice of a visual mode or of an audio mode. The visual mode will be shown in two ways: the first as a time series and the second as a matrix. (lower two plots of AVIPE). The RueBea team will coordinate operator interface requirements and configure the AVIPE GUI to address the operational needs of the intelligence gathering and analysis process.

1.2 Potential Impact to the Commercial and Non-commercial Entities.

Both of the Commercial and Non-commercial entities utilize various data processing methods to derive actionable intelligence from different information sources. Improved data processing methods are desired to increase efficiency and to extract new types of information from the vast amount of data that is being collected and analyzed in our present state of virtual non-stop data streams. AVIPE will provide a new capability to screen intelligence data that will result in a unique facilitator for the intelligence analyst. The AVIPE cepstral-based process in conjunction with the GV 3.0 GUI will be of benefit to the important tasks at hand by reducing the amount of manpower and effort devoted to less significant data and directing this valuable manpower and analyst asset to areas of information that warrant more attention from a Business-value perspective.

Building on Figure 1, Figure 2 demonstrates the relationship between the series of binary digits (bits, John Tukey circa 1977) flowing through the network card and the loading of a webpage onto the operators' monitor using the information in the (blue and red, 1's and 0's) bits. Initial research has determined that there detectable differences in the (underlying) traffic communication serving to build the page. AVIPE can determine the difference by observing 'the pulse' of the network traffic flow. For example, supposed that a keyword is an object of interest residing on the vertical yellowed section in the figure. The existing method of detection would be to transform the keyword to a bit level equivalent, continuously correlate and wait for a match. The problems with this type of analysis are that:

1. The word may be printed outside of the yellowed section causing false alarms.
2. A delay in sending the yellowed sectioned piece, thus wasting time correlating against an idle traffic pattern.

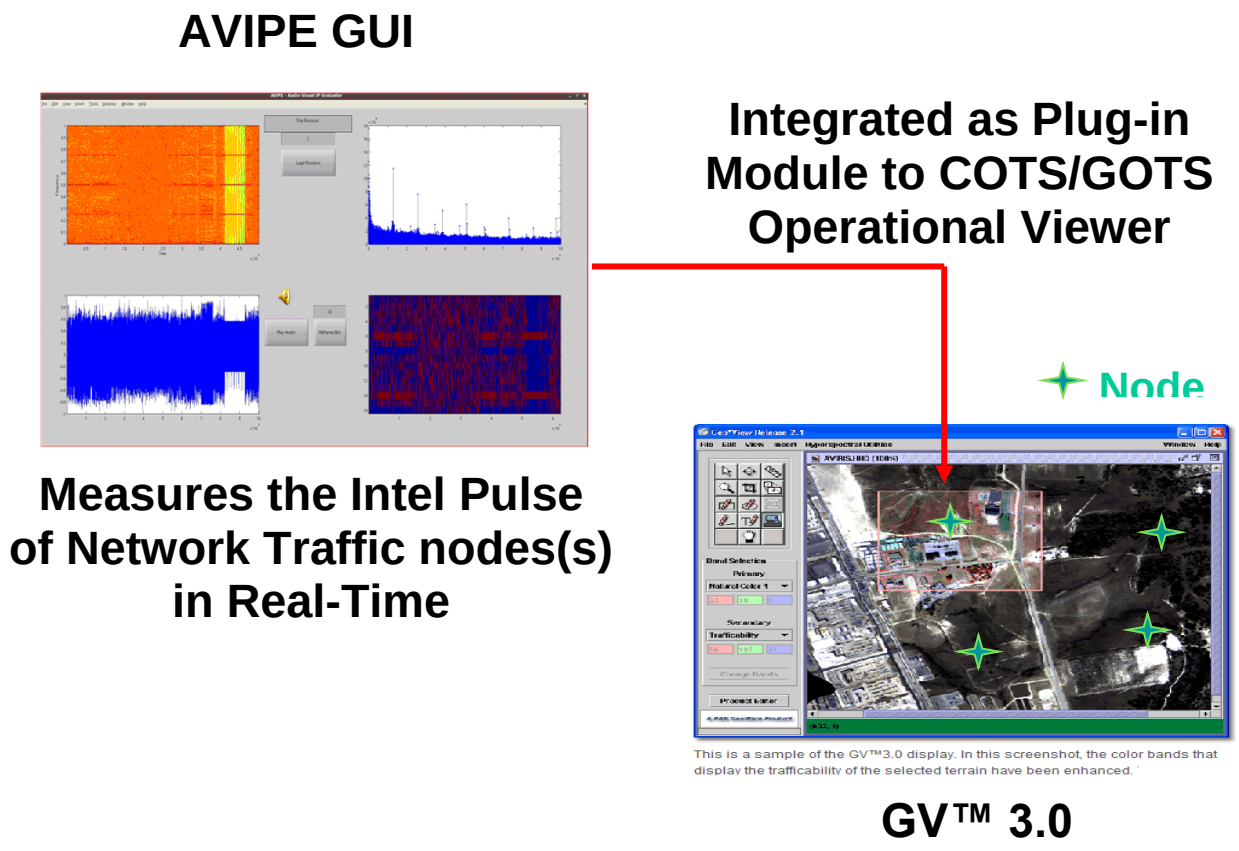

Figure 1 – AVIPE/GV 3.0 GUI Viewer – Assigning AVIPE analysis to particular set of nodes in the field.

AVIPE is a tool that would listen for or observe a discernible change in the pulse of the traffic as it started up the creation of the yellow banner which is then finished before any words are printed on it. Thus this technique would:

1. detect the building of the yellowed section separate from the remaining parts of the page.
2. Cut down on operator observation time and utilizes other (auditory) sense.
3. Provide observation matrix for image processing and shape analysis.
4. Facilitate 'course filter' awareness among nodes in sensor field.

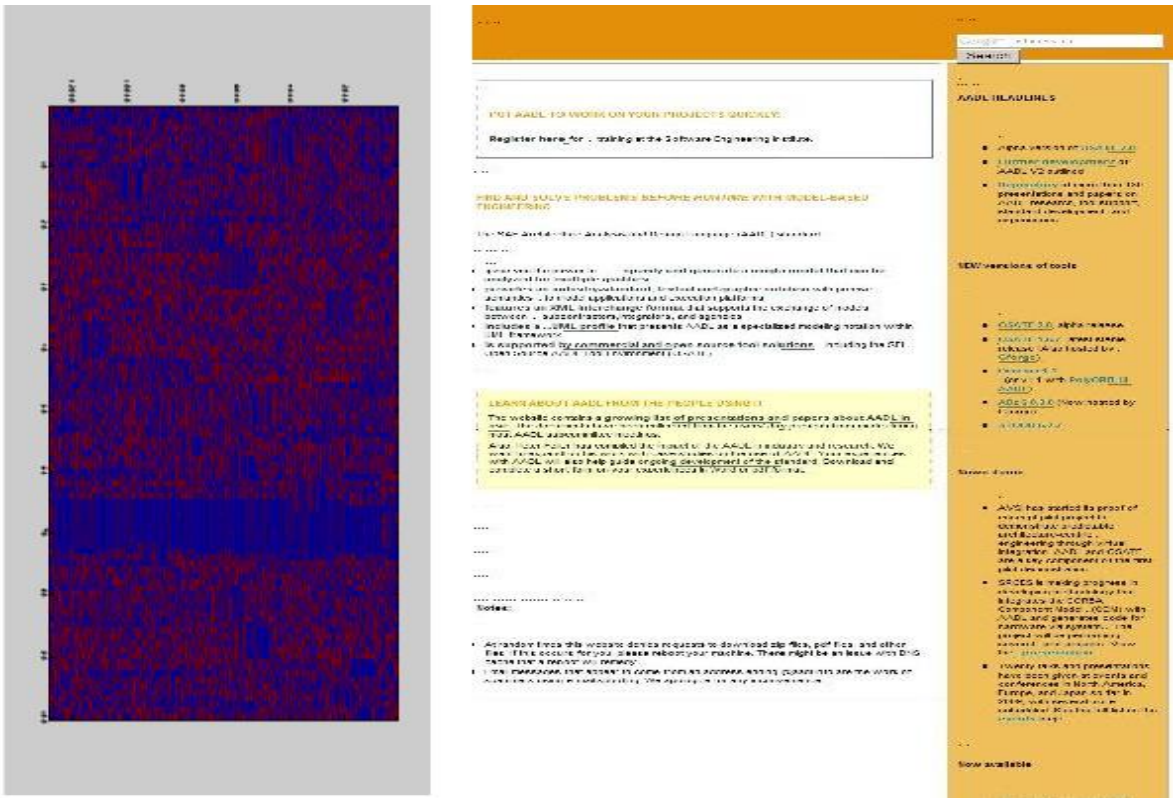

Figure 2 - Bit domain on left relates to sequential loading of web page. Cepstral coefficients associated with yellow (vertical) banner can queue when text is about to be written into banner.

1.3 Innovativeness of the Effort.

The proposed effort is innovative in its approach to use and apply cepstral processing techniques to detect and identify key elements of information contained in live network Intel data traffic. Cepstral processing is an accepted and recognized technology in the speech domain, but no known significant effort has been expended to Intel data in the IP domain. Currently, word or protocol matching is a typical method used to monitor network traffic. However, these matching methods are insufficient to meet any high data throughput demand and are more suitable for

addressing searches of archived information resources. AVIPE will work on the analysis of live dynamic IP data traffic. AVIPE will alleviate the processing burden such that matching techniques can be directed to and focused on suspected subsets of information. As with any coarse filter there will be some false alarms and there will be additional time required to drill down for specified cases, however the dramatic gain in access to data will more than compensate for the second stage signal processing. As the (AVIPE/Viewer) tool becomes better accepted by the operator the audio playback will allow for the operator to work on other tasks thus being more efficient; like driving a car and listening to the radio. The project recognizes that software engineers constructing network traffic frames are actually constructing a deterministic set of dialect and idiolect components and hence the traffic information should be analyzed as a speech signal.

1.4 Uniqueness of the Effort to the NRO.

There is potential for use of the AVIPE capability by both commercial and other non-commercial customers. Classifying the uniqueness of the cepstral coefficients will be necessary to support the data analysis requirements of the end-users.

In review, AVIPE is a novel approach that brings a practical solution to the data overload problem. With the exponential increase in information being stored or passed among sensors/networks a tool like AVIPE is needed to make a practical attempt at choosing coarse events that are homogeneous using some type of measure; in this case our measure relies on a modified cepstrum. AVIPE will be a prelude to other search tools. AVIPE can be thought of as a measure of a ***pulse*** of the network traffic, ***in real time.*** In a unique way the entities who embrace this approach will be the first able to listen to changes in the network traffic traveling at the speed of light by moving beyond the constraints and principles set forth by Nyquist and yet rendering an intelligible insight into how the traffic is behaving.